A Quick Start Guide to Survey Research

Liz Carey (and hopefully many others)

2019-01-17

Contents

Welcome to survey research

This book is intended to be a quick resource for conducting survey research. By no means is it intended to be comprehensive of all survey research methodologies.

Preface

Hopefully you'll find this book to be a condensed and easy to read resource on survey research.

We developed this book in the hopes of future collaboration among other UX researchers.

Outline

The content of the book will include:

- **Chapter 1**
- **Chapter 2**

Prerequisites

All you need is an interest in conducting survey research and basic data analysis, we'll include code snippets (python and R) along the way.

Acknowledgements

This book wouldn't be possible without the contributions of:

Chapter 1

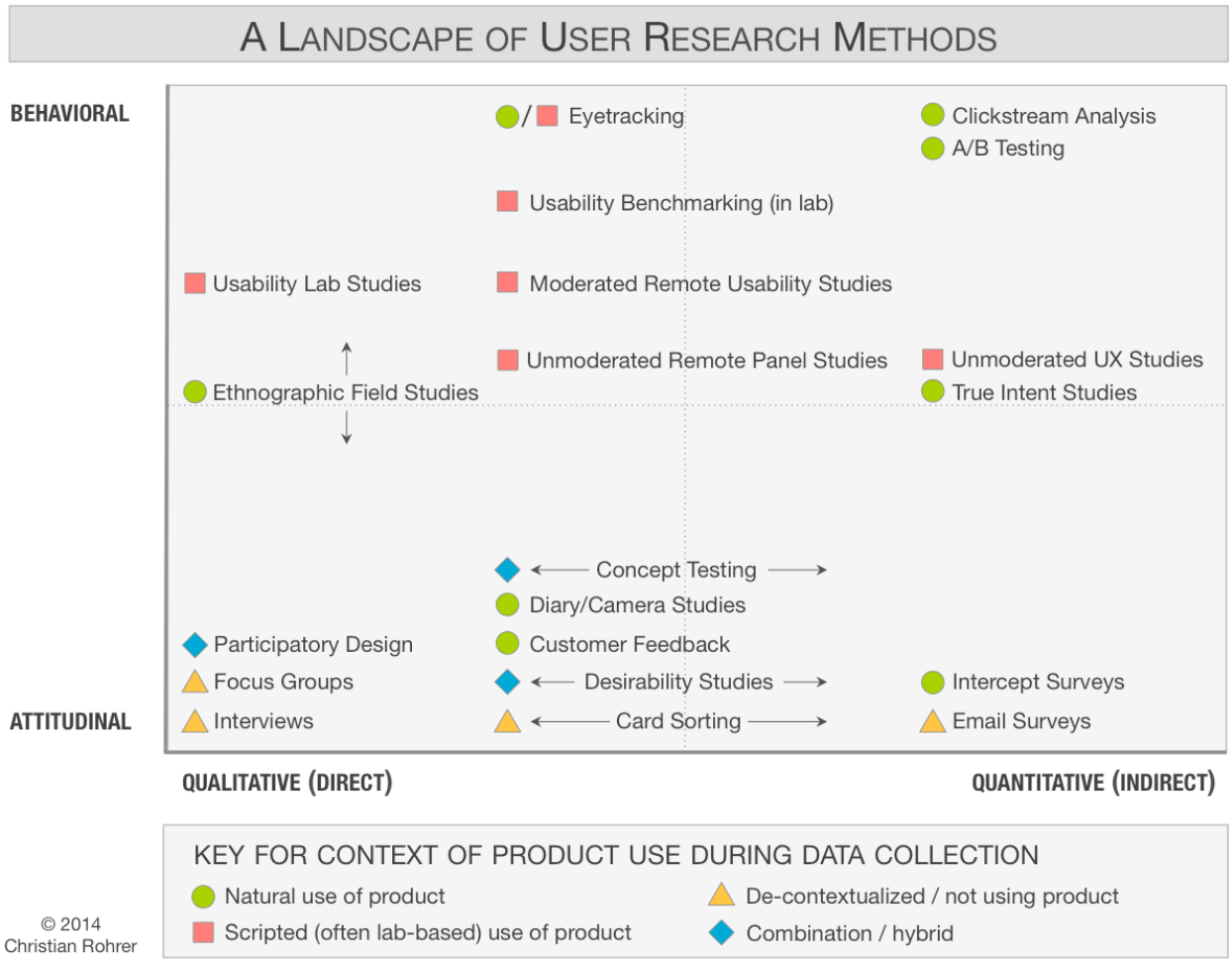
Designing a survey

1.1 What is your research goal?

First, establish if a survey is the right method to accomplish your research goal by asking yourself:

- What do you currently know?
- What *don't* you know?

Below is a useful visualization from the Nielsen Norman group on how to decide between which qualitative or quantitative methods to answer your research goal (Rohrer, 2014).



Surveys are great for answering the "How many and how much" of what people do and say; surveys are not the best method at understanding the "Why and how to fix" a product problem.

1.2 Who are you studying?

This question may be simple at first, but when you start to narrow down

Chapter 2

Writing effective survey questions

Effective survey questions result in **consistent** and **reliable** responses.

Chapter 3

Survey Analysis

After you've fielded your survey, here are the steps to making sense of the data.

This section assumes you have a laptop set up to work with in either R or python. Head over to the Appendix page if you need help with set up.

3.1 Organize your workspace

Before beginning any analysis, you'll want to set up a reproducible workflow. Below is an adapted suggestion on how to organize your workspace from Ben Marwick, Carl Boettiger, and Lincoln Mullen (Ben Marwick, 2018). Keeping your workspace organized is the best way for you and others to understand and reproduce your analysis.

```
project
|- DESCRIPTION          # project metadata and dependencies
|- README.md            # top-level description of content and guide to users
|
|- data/                # data files used
|  +- raw_data.csv      # data files in open formats such as TXT, CSV, TSV, etc.
|  +- cleaned_data.csv  # data files that have been cleaned, merged, etc that you'll use for survey anal
|
|- analysis/            # any programmatic code
|  +- my_report.Rmd     # R markdown file with narrative text interwoven with code chunks
|  +- makefile          # builds a PDF/HTML/DOCX file from the Rmd, code, and data files
|  +- scripts/          # code files (R, shell, etc.) used for data cleaning, analysis and visualisatio
|  +- figures/          # saved outputs of your figures
|
|- R/
|  +- my_functions.R    # custom R functions that are used more than once throughout the project
|
|- man/
|  +- my_functions.Rd   # documentation for the R functions (auto-generated when using devtools)
|
```

R version

```r
#List the directory names you want to create
folder_names <- c("data",
                  "data/raw",
```

```
                    "data/clean",
                  "analysis",
                    "analysis/scripts",
                     "analysis/figures",
                  "R")

#Create the directories
sapply(folder_names, dir.create)
```

3.2 Data Cleaning

Before you can begin looking at the results, you'll need to clean the data. By "cleaning" the data, we mean edited the raw file into a format that will make the analysis valid and easier.

3.2.1 Load the data

Download your raw survey data as a csv and load it into your your analysis tool of choice (e.g. Ipython notebook or Rstudio)

R version

```
#load necessary packages for analysis
library(tidyverse)        #contains all the library packages to manipulate and transform data
library(summarytools)     #shortcut tools to visualize summaries of the data

#read/store the data as the variable df (short for dataframe)
#replace "file" with "https://raw.githubusercontent.com/lizmcarey/survey-guide/master/sample_data/Surve
df <- read_csv(file)
```

python version

```
#load necessary modules for analysis
import pandas as pd

#read/store the data as the variable df (short for dataframe)
df = pd.read_csv(filename)
```

3.2.2 Loading Qualtrics data

When you download a csv from Qualtrics, it will come with a few extra rows you don't need. Here are some automated scripts you can add to your makefile to speed up your work =flow

R version manual

```
#Store the column names by reading in the column header
df_names <- read_csv(file, n_max=0) %>% names()

#Read the entire file,
df <- read_csv(file,
               col_names = df_names, #use df_names to title the columns
               skip = 3) #skip the first three lines
```

```
#store the question names
question_bank <- read_csv(file, n_max=1) %>%  #read in the first row of file
                 select(starts_with("Q")) %>% #select columns that start with Q
                 gather(key, question_text) #move data from wide to long
```

R version programmatic

```
load_qualtrics_csv <- function(file) {
  df_names <- read_csv(file, n_max = 0) %>% names()

  df <- read_csv(file, col_names = df_names, skip = 3)
}

get_questions <- function(file) {
  qb <- read_csv(file, n_max = 1) %>%
        select(starts_with("Q")) %>%
        gather(key, question_text)
}

df <- load_qualtrics_csv(file)

question_bank <- get_questions(file)
```

3.2.3 Preview the data

It's important to get a look at the data to spot an errors in uploading, etc.

Appendix A

Setting up R

A.1 Package installation

You'll want to install the following packages:

```
library(tidyverse)
```

Appendix B

Setting up python

```
# Pandas makes working with data tables easier
import pandas as pd

# Numpy is a library for working with Arrays
import numpy as np

# Module for plotting graphs
import matplotlib.pyplot as plt
import seaborn as sns

# SciPy implements many different numerical algorithms
import scipy.stats as stats
import collections
```

Bibliography

Ben Marwick, Carl Boettiger, L. M. (2018). Packaging data analytical work reproducibly using r (and friends). *PeerJ*.

Rohrer, C. (2014). When to use which user-experience research methods. *Nielsen Norman Group*.